



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV RADIOELEKTRONIKY

DEPARTMENT OF RADIO ELECTRONICS

AUTOMATICKÁ DETEKCE ŠTĚKAJÍCÍCH PSŮ PODLE ZVUKOVÉHO SIGNÁLU

AUTOMATIC DETECTION OF BARKING DOGS BASED ON SOUND SIGNAL ANALYSIS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Kamil Káčer

VEDOUCÍ PRÁCE

SUPERVISOR

prof. Ing. Milan Sigmund, CSc.

BRNO 2021

Bakalářská práce

bakalářský studijní program **Elektronika a komunikační technologie**

Ústav radioelektroniky

Student: Kamil Káčer

ID: 211777

Ročník: 3

Akademický rok: 2020/21

NÁZEV TÉMATU:

Automatická detekce štěkajících psů podle zvukového signálu

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s problematikou zpracování zvukových signálů. Zpracujte rešerši dostupných databází a významných publikací pojednávajících o analýze zvuků vydávaných zvířaty. Jednotlivé vybrané algoritmy (alespoň tři) na detekci psího štěkotu v časové a kmitočtové oblasti naprogramujte a ověřte na reálném signálu z existující databáze. Určete nutnou minimální délku zvukového signálu pro správnou činnost detekčních algoritmů.

Vytvořte databázi psího štěkotu s doprovodnými údaji o štěkajících psech (rasa, stáří, velikost). Databázi doplňte o štěkot simulovaný člověkem. Vytvořte ucelený program na spolehlivou detekci psího štěkotu v akusticky různorodém signálu. Vyhodnoťte, které jiné zvuky váš program rozpoznává jako štěkot. Výsledný software v jednom provedení optimalizujte na detekci vybraného psího jedince. Spolehlivost detekce statisticky vyhodnoťte.

DOPORUČENÁ LITERATURA:

[1] JAN, J. Číslíková filtrace, analýza a restaurace signálů. Brno: VUT v Brně, 2002.

[2] GIANNAKOPOULOS, T., PIKRAKIS, A. Introduction to Audio Analysis: A MATLAB Approach. New York: Academic Press, 2014.

Termín zadání: 8.2.2021

Termín odevzdání: 27.5.2021

Vedoucí práce: prof. Ing. Milan Sigmund, CSc.

prof. Ing. Tomáš Kratochvíl, Ph.D.
předseda rady studijního programu

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

ABSTRAKT

Táto práca sa zaoberá detekciou psieho štekotu spomedzi iných zvukov. Tak isto práca obsahuje niektoré postupy iných autorov a je určená minimálna dĺžka psieho štekotu pre správnu detekciu. Postupy pracujú s príznakmi z kmitočtovej aj časovej oblasti a výsledky sú vyhodnotené.

KLÍČOVÁ SLOVA

signál, príznaky, pes, detekcia, frekvencia

ABSTRACT

This semestral thesis presents detection of barking dogs amongs other sounds. This thesis deals with approaches that authors have used in the past. Further more the minimum length of a dog bark was estimated. Parameters worked in both frequency and time domain and the results are shown and discussed at the end.

KEYWORDS

signal, parameters, dog, detection, frequency

KÁČER, Kamil. *Automatická detekce štěkajících psů podle zvukového signálu*. Brno, 203045, 40 s. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: prof. Ing. Milan Sigmund, CSc.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Automatická detekce štěkajících psů podle zvukového signálu“ jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

podpis autora

PODĚKOVÁNÍ

Rád by som sa poďakoval vedúcemu semestrálnej práce pánovi prof.Ing.Milanovi Sigmundovi, CSc za konzultácie a veľkú trpezlivosť.

Obsah

Úvod	9
1 Základné poznatky o signáloch	10
1.1 Signály	10
1.2 Akustický signal	10
1.3 Periodicita signálu	10
1.4 Počuteľné pásmo	11
1.5 Frekvenčné spektrum signálu	12
1.6 Časté formáty zvukových súborov	13
2 Metódy na analýzu zvieracích zvukov	15
3 Metódy pre detekciu psieho štekotu podľa zvukového signálu	17
3.1 Audio príznaky	17
3.2 Metódy detekcie	19
4 Vytvorenie databázy	25
5 Vytvorenie uceleného programu na spoľahlivú detekciu	27
5.1 Testovanie	29
5.2 Určenie Threshold	30
5.3 Vyhodnotenie	31
5.4 Program Môj Pes	32
5.5 Princíp a inštrukcie ku práci s programom	34
Záver	36
Literatura	38
Seznam symbolů, veličin a zkratek	40

Seznam obrázků

1.1	Zobrazenie zvuky v časovej doméne	10
1.2	Zvukové rozsahy ktoré su živočíchy schopné vnímať	11
1.3	Logaritmicke vnímanie zvuku u človeka	12
1.4	Spektrum signálu (Psí brechot)	12
1.5	Diskreditácia analógového signálu	13
3.1	Úkažka príznaku zerocrossing rate	17
3.2	Princíp oknovania	19
3.3	Popis zapisovania do matice	20
3.4	Stavový diagram prvého algoritmu	21
3.5	Vykreslenie spektier pre Vzor a Vzorku	22
3.6	Porovnanie spektra výstrelu a štekotu	23
3.7	Vykreslenie pre vzor a vzorku Mel frekvenčných koeficientov	24
5.1	Príklady dvoch rôznych štekotov a ich koeficienty	29
5.2	Porovnanie koeficientov	30
5.3	Výsledné zobrazenie	31
5.4	Spektrum zvuku vydávané sliepky	33
5.5	Spektrum psieho štekotu	34
5.6	Zjednodušený stavový diagram pre Rozpoznávanie psích štekotov	35

Seznam tabulek

4.1	Prehľad zvukov v databáze	25
5.1	Koeficienty vzorovej matice pre malého psa	27
5.2	Koeficienty vzorovej matice pre veľkého psa	28
5.3	Výsledná pravdepodobnosť, hodnoty Treshold a počet ponechaných koeficientov	32

Úvod

Tato práca sa venuje oblasti DSP (Číslicové zpracování signálů – Digital Signal Processing), Konkrétne analýzy audio segmentov psieho štekotu z ktorého sú následne extrahované príznaky. Celkový softvér je na koniec optimalizovaný pre jedného jedinca a spoľahlivosť detekcie je vyhodnotená.

Pod pojme štekanie rozumieme aj rozličné zvuky ktoré sprevádzajú tento jav ako napríklad vytie alebo vrčanie.

Štekot sa prenáša vzduchom pomocou akustických vln v akustickom prostredí v pásme ktoré je počuteľné pre ľudí.

Súčasne sa touto oblasťou spracovania a detekcie zvukových signálov zaoberá stála viac vedcov z dôvodu širokého praktického využitia ako napríklad Audio klasifikácia rozpoznávanie reči či MIR(Music information retrieval).

Cieľom tejto práca je vytvoriť tri postupy na spoľahlivú detekciu, či už to v časovej alebo spektrálnej oblasti, pre určenie psa z spomedzi iných zvukov.

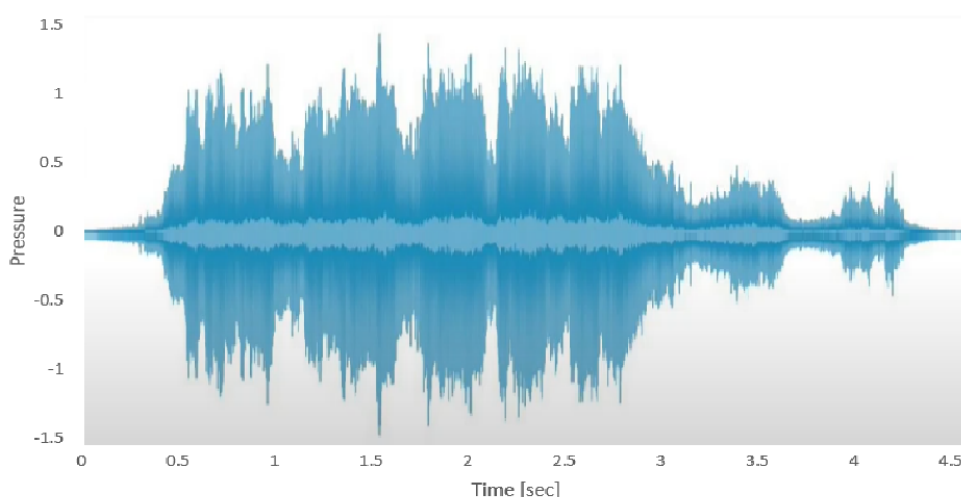
1 Základné poznatky o signáloch

1.1 Signály

Signál je fyzikálna alebo chemická veličina ktorá nesie určitý druh informácie. Táto informácia je buď už v signále obsiahnutá a užívateľ ju pozoruje, alebo je vložená samotným užívateľom na základe už stanovených pravidiel.

1.2 Akustický signal

Poznáme viaceré typy signálov napríklad akustický obrazový alebo optický. Signál ktorý je pre nás najviac zaujíma je akustický. Akustický signál alebo zvuk, je spôsobený nejakým predmetom, ktorý vydáva vibráciu. Táto vibrácia spôsobuje že molekuly zvuku začnú oscilovať a narážať do seba čo spôsobuje zmenu tlaku vzduchu a tým vzniká vzduchová vlna ktorá prenáša energiu z miesta na miesto.



Obr. 1.1: Zobrazenie zvuky v časovej doméne

Vo všeobecnosti zvuková vlna ako na Obr. 1.1 nesie informáciu o frekvencii intenzite a farbe daného zvuku v podobe uvedeného grafu.

1.3 Periodicita signálu

Všetky zvukové vlny rozdeľujeme do dvoch kategórii podľa ich periodicity na periodické a aperiodické. Periodické signály môžeme ešte rozdeliť na jednoduché a komplexné a aperiodické signály na kontinuálne a tranzientné.

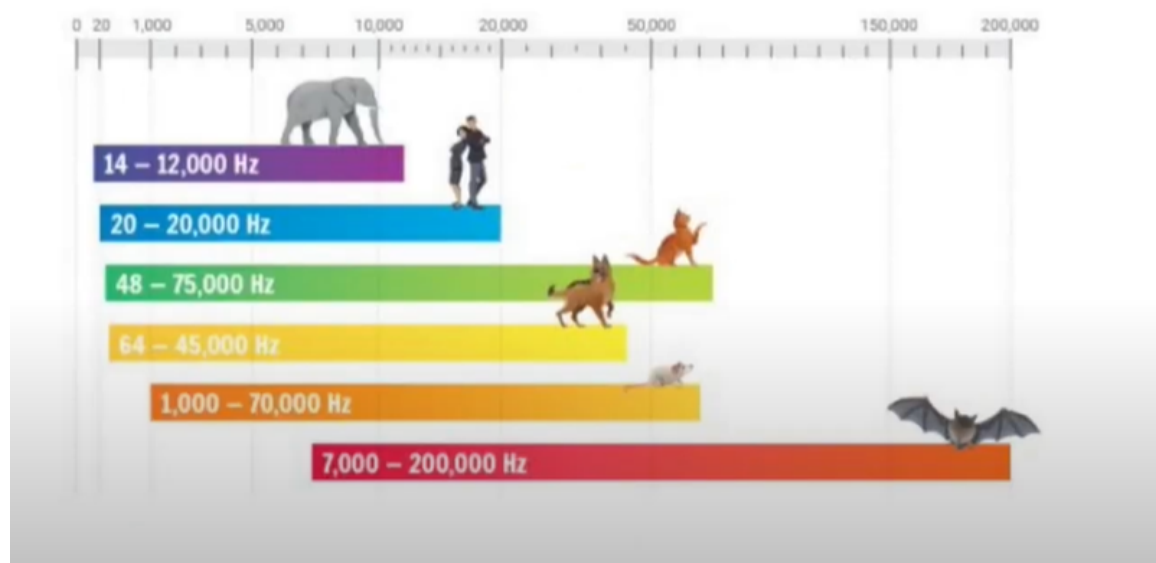
Medzi jednoduché periodické signály patrí napríklad funkcia sinus kdežto u komplexných nájdeme zvuky ako napríklad zhuk sínusových signálov z hudobných nástrojov na koncerte.

Pod kontinuálne aperiodické signály spadajú šum čo je vlastne nejaký náhodný a nepredvídateľný jav alebo zmena v akustickom tlaku. Tranzientné signale si môžeme predstaviť ako pulz čo je vlastne rýchla a výrazná zmena v akustickom tlaku. Ako zástupcu si môžeme spomenúť psí štekot alebo výstrel zo zbrane.

1.4 Počuteľné pásmo

Rozsahy frekvencií, ktoré sú počuteľné sa menia v závislosti na človeku alebo inej bytosti ktoré vníma zvukovú vlnu. Ako je vidno na obrázku Obr. 1.2 človek dokáže vnímať frekvencie ktoré sú v rozsahu od 20-20000Hz(dôležité poznať tieto hodnoty pretože od toho sa odvíjajú vzorkovacie periódy frekvencie atď)

Zvuk ktorý presiahne frekvenciu 20000Hz a ktorý je počuteľný napríklad psami alebo netopiermi sa nazýva ultrazvuk a zvuk ktorý počujú napríklad slony a je pod frekvenciou 20Hz nazývame infrazvuk.



Obr. 1.2: Zvukové rozsahy ktoré su živočíchy schopné vnímať

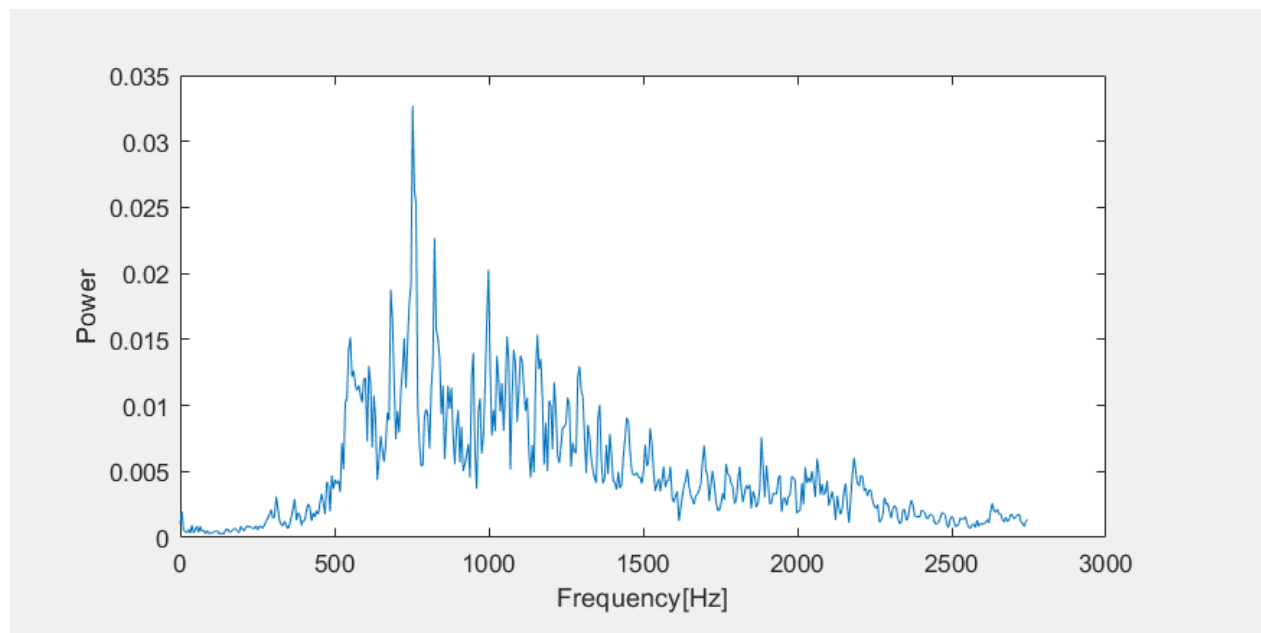
Ďalej je vhodné spomenúť že človek nevníma frekvenciu ani amplitúdu zvuku lineárne ale logaritmicky ako ukazuje Obr. 1.3 Tento fakt si je dobre uvedomiť pretože niektoré parametre ako napríklad MFCCs(Mel Frekvenčné Cepstrálne koeficienty) pracujú s touto myšlienkou.



Obr. 1.3: Logaritmické vnímanie zvuku u človeka

1.5 Frekvenčné spektrum signálu

Frekvenčné spektrum signálu je zobrazenie určitých frekvenčných zložiek ktoré majú na výsledný priebeh najväčší vplyv. Ako demonštruje Obr. 1.4 tak zhruba najväčší vplyv na signál majú zložky okolo 700Hz.



Obr. 1.4: Spektrum signálu (Pší brechot)

Prevod z časovej domény Obr. 1.1 do frekvenčnej Obr. 1.4 sa realizuje pomocou Fourierovej transformácie a najefektívnejšie pomocou FFT (Fast Fourier Transform) ktorá sa využíva najviac keďže je najrýchlejšie spomedzi všetkých algoritmov, ktoré sa používajú na tento časovo-frekvenčný prevod.

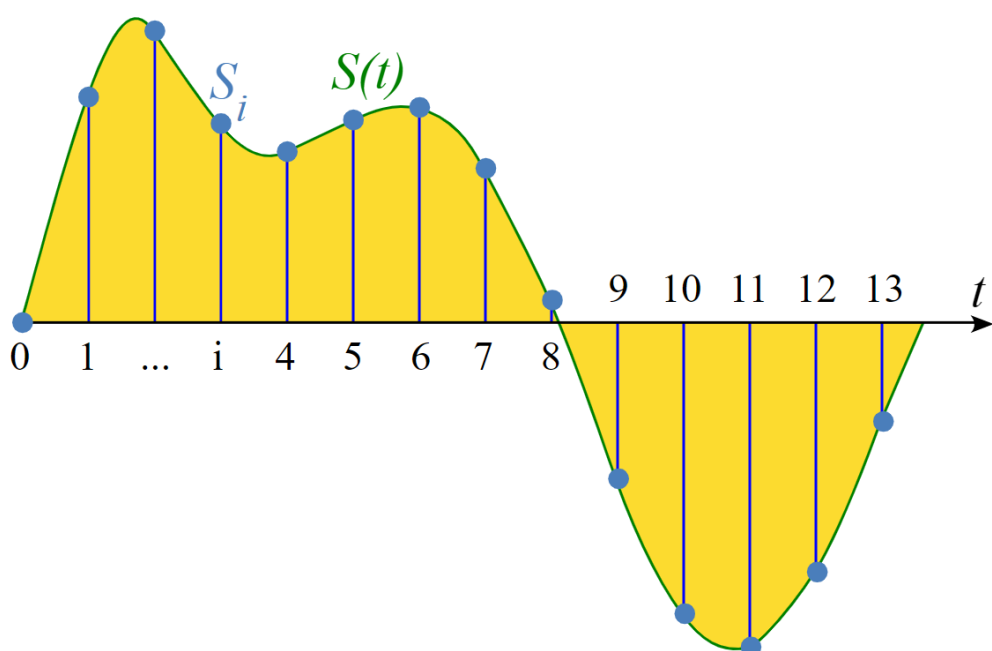
1.6 Časté formáty zvukových súborov

Pre ukladanie zvukových súborov sa používajú viaceré formáty. Zvukový formát je popis zvuku akým spôsobom bol uložený do súboru.

Pred tým ako si spomenieme niektoré najčastejšie zvukové formáty je dobré poznať pojmy ako vzorkovacia frekvencia alebo bitová hĺbka. Príklad vzorkovacej frekvencie môže byť 44,1kHz. Keď sa použije táto hodnota na zvukový signál v časovej doméne znamená to že každú jednu sekundu je signál preseknutý 44100 krát a v každom preseknutom bode sa vytvára vzorka. Vzorkovacia frekvencia môže byť samozrejme aj vyššia. Bitová hĺbka je číslo ktoré keď dosadíme do vzorca za n

$$m = n^2$$

dostávame počet kvantizačných úrovní ku ktorým môžu byť nevzorkované vzorky priradené. Príklad bitovej hĺbky 16 ,24.



Obr. 1.5: Diskreditácia analógového signálu

Zvukové súbory rozdeľujeme do dvoch hlavných kategórií na bezstratové formáty a stratové formáty. Medzi najznámejšie a najpoužívanéjšie stratové formáty patria

MP3 alebo ACC. Stratovosť sa prejavuje redundanciou čo je vlastne odstraňovanie nepotrebných zvukových informácií na základe psychoanalytického modelu. Ide o zvuky ktoré človek nie je schopný oceniť alebo nie sú podstatnou súčasťou zvukovej nahrávky. Spomenuté formáty sú často používané pre ich malé veľkosti. Ide o horšiu kvalitu zaznamenania zvuku.

Ku najviac používaným bezstratovým nekomprimovaným zvukovým formátom patria formáty ako Wav (Waveform audio file formát) alebo Flac (Free Lossless Audio Codec).

Wav format

V tejto práci sa pracuje výhradne s formátom wav, ktorý bol vyvinutý firmami IBM a Microsoft. Ide o nekomprimovaný audio formát kódovaný LPCM (Linear Pulse code modulation) vzorkovaný 44100 Hz s bitovou hĺbkou 16bit na jednu vzorku. Ide o formát s vysokou kvalitou a dostupnosťou a preto bol vybraný pre túto prácu.

2 Metódy na analýzu zvieracích zvukov

V tejto kapitole bude spomenuté niekoľko postupov a stručný prehľad najlepších a najpoužívanejších metód na analýzu zvukového signálu u zvierat najmä u psov, ktoré použili a sú používané vedcami v súčasnosti.

Práca *Barking in domestic dogs: context specificity and individual identification* od autorov Sophie Yin a Brandy McCowan sa zaujímala či psí štekot môže byť rozdelený na základe kontextu pri akom boli štekoty zaznamenané. (vyrušenie, hranie, izolácia). Dáta boli zozbierané počas necelého pol roka pre niekoľko rozličných plemien a pri spomenutých situáciach. Dáta boli nahraté a digitalizované pri 16bitoch a pri vzorkovacej frekvencii 44100Hz. Dĺžka jedného štekotu bola medzi 38-137ms. Analýza pozostávala so 4672 vzoriek

Pre kategorizáciu jednotlivých štekotov bola použitá DFA (Discriminant function analysis). Táto multivariabilná analýza sa používa keď jeden príznak sa mení v užitočnosti. Napríklad môže byť jednoduché rozpoznať prvý štekot od druhého na základe priemeru frekvenčných zložiek ale použitie toho istého príznaku nemusí stačiť u prvého a tretieho štekotu. Tam sa zase ako obmena použijú príznaky ako dĺžka štekotu a priemer amplitúdy a meranie prebehne presnejšie. Multi variantná analýza kompenzuje tieto nedostatky. Na začiatku sa určia, ktoré akustické hodnoty (Rozsah amplitúdy, Minimálna frekvencia, atď) majú najväčší podiel na rozdeľovaní a ktoré sa ďalej použijú v Cross validation discriminant function analýze na kategorizovanie daného štekotu. Touto metódou dosiahli autori nasledovné výsledky. 80% pre štekot vo vyrušení 75% štekot v izolácii a 75% pre štekot pri hraní. V tejto metóde bola použitá sporadicky aj Univariabilná analýza, ktorá používa vopred stanovených nemenných príznakov z dôvodu menšej presnosti bola používaná viac metóda Multi-variabilná.

V ďalšej práci s názvom *Detection of continuous barking actions from search and rescue dogs* ktorej autormi Yuichi Komori, Kazuaki Ohno a Takuaki Fujieda sa zaoberali lokalizovaným obetím zapadnutými pod snehom pomocou psích štekotov. Ich tvrdenie je, že pomocou kontinuálneho štekanie sa dá určiť kde sa obeť nachádza. V tomto algoritme bol použitý iba jeden príznak a tým bol tvar signálu. Nerozlišovalo sa medzi rôznymi druhmi psov iba im šlo o to, aby pri dlhom úseku nahrávania v reálnom čase bolo možné určiť kedy pes šteká od iných zvukov ako napríklad šum let helikoptéry. a aby následné mohli lokalizovať psa pomocou aj vibračného zariadenia ktoré taktiež bolo pripevnené k psovi. Keďže išlo o nie náročný postup dosahovali sa dobré výsledky cez 90%.

Pre klasifikovanie bola použitá metóda DTW (Dynamic time warping) táto metóda ráta korelačnú vzdialenosť medzi dvoma signálmi.

Dáta boli zozbierané pomocou Cyber suit ktorí mal pes pripevnený na sebe počas

hľadania obete. Na začiatku tri trérovacie sekcie boli uskutočnené keď sa obeť ukryli v teréne a čakali kým ich pes nájde a následne zašteká. Ako príznak bol použitý signály tvar.

Posledná spomenutá práca od autorov Che Yong Yeo a Chee Kyun Ng s názvom 'Dog Voice Identification (ID) for Detection System' sa zaoberá vývojom algoritmu na detekciu zvukových zvieracích vzorov. Ako príznaky použité v tejto práci sú Zero crossing rate, ktorý detekuje koľko krát sa signál pretne s časovou osou a ktorí sa využíva na detekciu ticha - príznak MFCC. Ako klasifikátor je použitá metóda Dynamic Time Wrapping DTW ktorá spája vloženú vzorku s natrénovanou v celej databáze.

Prvý krok potom čo je signál privedený na vstup je zistiť kedy presne začína a končí zvuk - odstraňujú sa tiché miesta a ďalej sa pracuje čisto so zvukom bez tichých miest. Ďalej je na rade príznaková extrakcia kde sa použije príznak MFCC kde ako výstup dostávame kompaktný zvukový model ktorý sa lepšie analyzuje (nepotrebné časti sa vynechávajú). Extrakcia prebehne v každej časti audio segmentu na ktoré je celý štekot rozsekaný.

Pre klasifikáciu sa používa metóda DTW ktorou podmienkou je aby obidva signály boli časovo zladené následne sú dva signály vložené do matice a DTW určí cestu v tejto matici ktorá najlepšie reprezentuje podobnosť vzoriek v testovacom a referenčnom signáli.

Nový Mat súbor je generovaný vždy keď je nový zvuk uložený do databázy.

Nakoniec sa vykoná spomínaná DTW metóda ktorá určí vzdialenosť medzi vloženým zvukom a zvukmi obsiahnutými v databáze a ku ktorému zvuku má vložený zvuk najbližšie tak je vyhodnotený ako ten zvuk.

Ako referencia alebo vzory boli použité 4 zvuky. Úspešnosť klasifikácie prvej referencie bolo 70% úspešnosť druhej do druhej referencie bolo 80 % úspešnosť do tretej 70% a do poslednej 90%. Pracovalo sa všetko so súbormi wav.

Databázy

Výber databáz závisí na veľa faktoroch avšak pre audio klasifikáciu a v tejto práci sú najvhodnejšie databázy GTZAN, databázy so stránok Labrosa alebo Berlin Database of emotional speech. Niektoré sú voľne prístupné a niektoré vyžadujú poplatok za prístupenie.

3 Metódy pre detekciu psieho štekotu podľa zvukového signálu

3.1 Audio príznaky

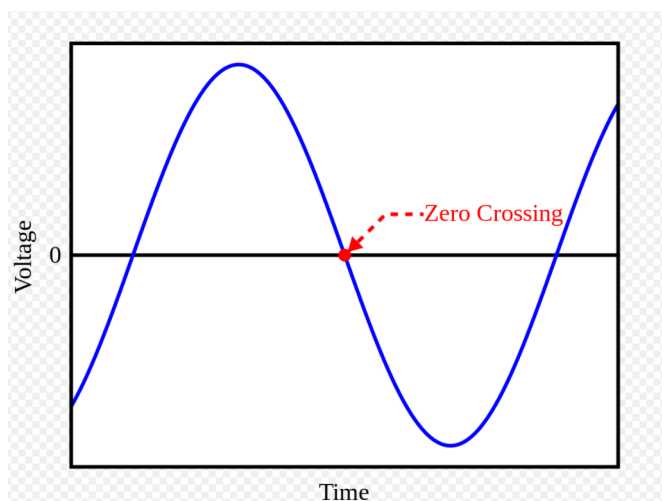
Termín audio príznaky je veľmi dôležitý termín v oblasti detekcie audio signálu jedná sa o vlastnosť ktorá opisuje charakter zvuku. Odlišné príznaky zachytávajú odlišné vlastnosti zvuku. Následne vybrané príznaky sú posunuté do strojového učenia, ktoré analyzuje jednotlivé podobnosti a vyriešia danú problematiku. Audio príznaky sa rozdeľujú na príznaky v časovej oblasti a na príznaky v oblasti frekvenčnej. V tejto práci sa používajú príznaky v časovej aj frekvenčnej oblasti.

Počet prechodu osou y Zero crossing rate

Jedná sa o príznak v časovej oblasti, ktorý analyzuje zadaný audio úsek výstupom ktorého je rýchlosť akou signál prechádza z pozitívnej časti na osi y do zápornej oblasti. Vzorec pre výpočet je

$$Z(i) = \frac{1}{2W_L} \sum_{n=1}^{W_L} | \operatorname{sgn}[x_i(n)] - \operatorname{sgn}[x_i(n-1)] | \quad (3.1)$$

Kde W_L je dĺžka jedného časového úseku.



Obr. 3.1: Úkažka príznaku zerocrossing rate

Ďalším príznakom, ktorý je použitý a taktiež patrí do časovej oblasti je príznak Energia. Príznak energia poskytuje takzvanú silu audio segmentu. Vzorec pre výpočet energie je

$$E(i) = \frac{1}{W_L} \sum_{n=1}^{W_L} |x_i(n)|^2 \quad (3.2)$$

Kde W_L je taktiež dĺžka jedného časového úseku.

Posledným použitým príznakom v časovej oblasti je príznak nazvaný Entropia energie. Entropia energie je interpretovaná ako miera náhlej zmeny levelu energie v audio signále. Ráta sa tak že každý malý úsek audio segmentu sa rozdelí na ešte menšie pod úseky, v ktorých je vypočítaná energia a následne tento úsek je vydelený celkovou energiou. Tento príznak sa používa pri signáloch s veľkou energiou, napríklad u výstrelu alebo psieho štekotu.

Sekvencia pravdepodobností sa vypočíta

$$e_j = \frac{E_{subFrame_j}}{E_{shortFrame_i}} \quad (3.3)$$

kde

$$E_{shortFrame_i} = \sum_{k=1}^K E_{subFrame_k} \quad (3.4)$$

Nakoniec sa entropia energie vypočíta pomocou vzorca

$$H(i) = - \sum_{j=1}^K e_j \cdot \log_2(e_j) \quad (3.5)$$

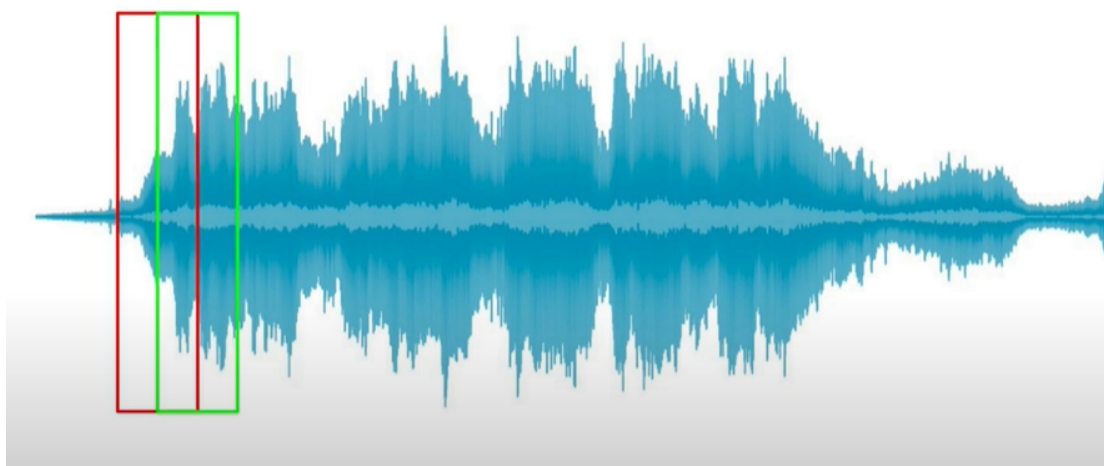
Ďalším použitým príznakom tento krát z frekvenčnej oblasti je príznak zvaný MFCCs alebo Mel-frekvenčné-cepstrálne koeficienty. Ide o naj-sofistikovanejší príznak použitý v tejto práci. MFCCs je cepstrálna reprezentácia signálu kde pri použití viacerých krokov ktoré upravujú amplitúdovú aj frekvenčnú oblasť vzniká nám ľuďom prijateľná miera. Používa sa logaritmus na amplitúdu a mel škálovanie na frekvenciu, kde obidve spomenuté domény prechádzajú z lineárnej reprezentácie na reprezentáciu logaritmickú prijateľnú pre človeka pretože ako bolo spomenuté človek vníma frekvenciu aj hlasitosť logaritmicky. MFCCs a ich extrakcia sa zameriava na farbu zvuku ale aj fonémy(najmenšie zvukové jednotka jazyka). Pri každej extrakcii je ako výstup najčastejšie 12 alebo 13 koeficientov ktoré najdôveryhodnejšie popisujú spomenuté zvukové javy.

3.2 Metódy detekcie

Prvý postup

Prvý algoritmus, je použitý v tejto práci je zobrazený v diagrame č. 3.4 Pred vykonaním postupu popísaného v spomenutom diagrame bolo rozhodnuté, že vzor nebude spriemerovaní z viacerých štekotov ale subjektívne sa vyberie jeden, ktorí najhodnovernejšieho reprezentuje štekot psa a ku nemu sa budú porovnávať iné zvuky.

Vo vzorke ako je nazvaná skupiny rôznych zvukov v diagrame č. 3.4 sú zvuky ako hra na klavíri, zvuk orchestru, hovorená reč, zvuky vydávané mačkou, šum a hra na flautu. Ako je ukázané na obrázku č.3.2 každé jedno okno, čo je na obrázku znázornené buď červenou alebo zelenou farbou, je vzor porovnávaný so vzorkou. Vzor sa posúva v dopredu určený časový interval tak aby nastalo prekrytie s predošlým audio segmentom. Po každom posunutí sa spočíta vzdialenosť medzi vzorom a vzorkou, podľa toho akú vzdialenosť užívateľ zvolí, buď Euklidovskú alebo Mahalanobisovu, sa výsledná vzdialenosť zapíše do bunky o počte, ktorí sa rovná počtu časových posunov okna. Jeden bunka obsahuje stĺpec o troch riadkoch v ktorých je uvedená vzdialenosť medzi vzorom testovacou a vzorkou.



Obr. 3.2: Princíp oknovania

Následné sa zoberú prvé prvky zo všetkých buniek a porovnávajú sa so zvolenou Threshold rozdielnou pre obidve vzdialenosti, pretože použitím špecifickej vzdialenosti dostávame iné hodnoty. Tento proces sa vykoná pre druhé aj tretie prvky v každej bunke (pre príznak 2 a 3) a výsledky sú zapísané do matice s názvom Result

Do matice Result sa zapisujú jednotky a nuly podľa toho či v danom audio segmente hodnota prekročila alebo neprekročila nastavenú hodnotu Threshold.

Hodnoty Threshold boli nastavené podľa testovania niekoľkých zvukov pre rôzne vzdialenosti (euklidovská mahalanobisova). Pre prvý príznak Zero Crossing rate boli nastavené Threshold na hodnoty 0.02 pre euklidovskú vzdialenosť a 0.08 pre mahalanobisovu. Pre druhý príznak Energiu bolo Threshold nastavené na 0.04 a 0.06 a pre posledný príznak Energiu Entropiu boli Threshold nastavené na hodnoty 0.25 pre euklidovskú vzdialenosť a 0.15 pre mahalanobisovu vzdialenosť.

Ďalej, ako je ukázané na Obr. 3.3 sa prechádza k postupnej horizontálnej a vertikálnej analýze matice. Pre každý príznak v stĺpci sa detekuje jednotka a ak aspoň jedna je prítomná v stĺpci, algoritmus vyhodnotí že štekot bol nájdený. Tento postup sa opakuje pre každý jeden stĺpec - každý príznak. Súčasne sa vykonáva aj vertikálna analýza, dôvodom tejto analýzy je určiť v ktorom audio segmente (frame) bol štekot detekovaný a tým pádom sa určí presný čas od začiatku celej testovacej vzorky kedy bol prítomný štekot.

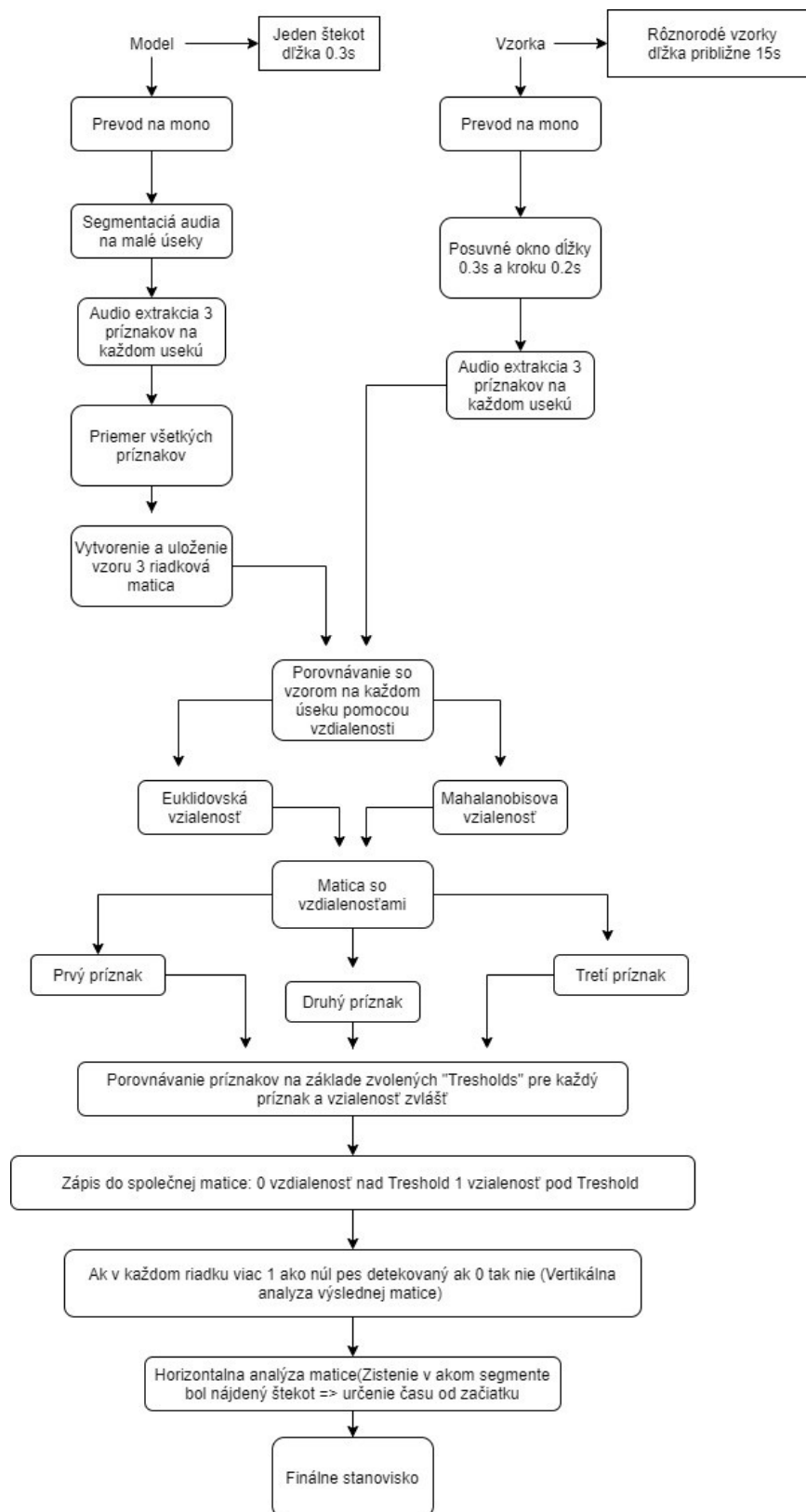
Na začiatku práce algoritmus pracoval len s horizontálnou analýzou matice, neskôr bola pridaná vertikálna ktorá určila presný čas kedy daný štekot nastal a tým pádom vylepšila celý postup.

Za najväčší problém spomenutého algoritmu je nastavenie presných hodnôt Threshold. Tým že je potrebné nastaviť šesť hodnôt dochádza v niektorých prípadoch k nepresnostiam. Extrahované príznaky nie sú presne stavané na štekot psa a preto podobné hodnoty sa pripisujú aj niektorým iným náhlým zvuk, ktoré v konečnom zúčtovaní majú podobné hodnoty ako štekoty psov. Táto chyba je čiastočne odstránená tým že sa nepoužíva jeden príznak ale viaceré - algoritmus sa spresňuje.

Číslo audio segmentu	Result			Horizontal_results	Display
	Prvý príznak	Druhý príznak	Tretí príznak		
1	0	0	0	0	
2	1	0	0	1	Dog detected frame 2
3	0	0	1	1	Dog detected frame 3
4	1	0	0	1	Dog detected frame 4
5	1	0	1	1	Dog detected frame 5
6	0	0	0	0	
7	0	0	0	0	
8	0	0	0	0	
9	0	0	1	1	Dog detected frame 9
Vertical_result	Dog_detected	0	Dog_detected		

Obr. 3.3: Popis zapisovania do matice

Pre prehľadný popis krokov 3.4



Obr. 3.4: Stavový diagram prvého algoritmu

Druhý postup

Druhý postup je založený na porovnávaní frekvenčných spektier. V tomto prípade na začiatku vytvorený s piatych psích štekotov. Každý jeden štekot bol pomocou FFT prevedený na spektrálne koeficienty a následne absolútnou hodnotou boli ponechané len reálne koeficienty, ktoré sú pre tento postup dôležité. Ďalej sa vypočítal priemer a bol stanovený vzor, ktorí sa vykreslil. V ďalšom kroku si užívateľ mohol vybrať zvuk, s ktorého taktiež pomocou FFT boli extrahované frekvenčné koeficienty a následne bola použitá korelácia na stanovenie podobnosti medzi maticami.

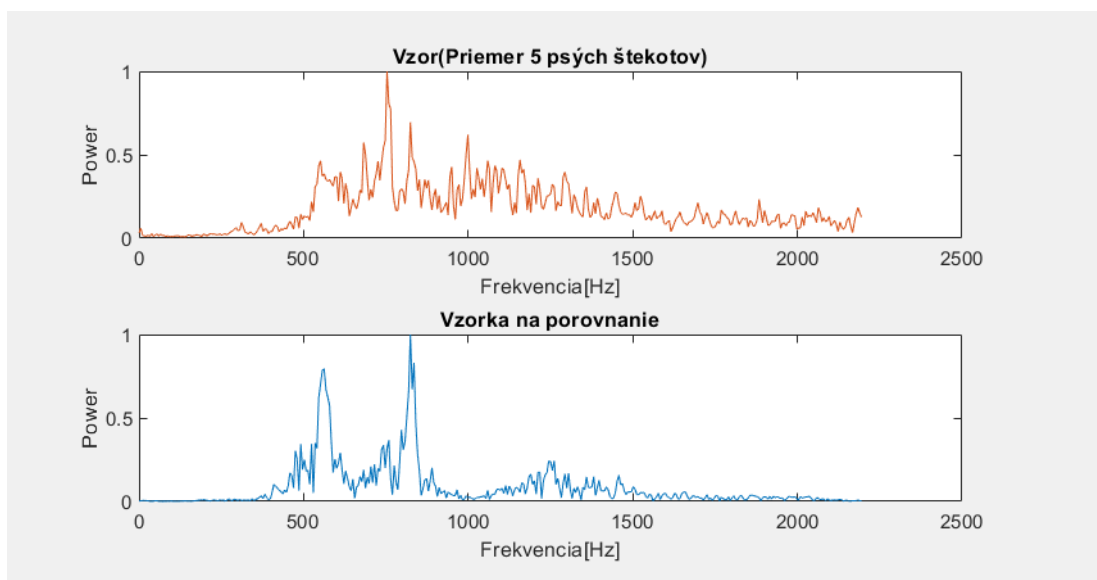
Ako v predošlom prípade, tak aj v tomto bola stanovená Threshold nad ktorú sa testovacia vzorka klasifikovala ako pes a pod ako iný zvuk.

V tejto metóde ja názorne ukázané ktoré frekvenčné zložky sa najviac podieľajú na spektre, majú najväčšiu silu. Ako je zrejmé z obrázka Č.3.5 tak na spektrum psieho štekotu majú najväčší vplyv zložky okolo 750Hz až 850Hz.

Súčasťou tohoto postupu malo byť aj paralelné porovnávanie Cepstrálnych koeficientov ktoré sa vypočítajú podľa vzťahu

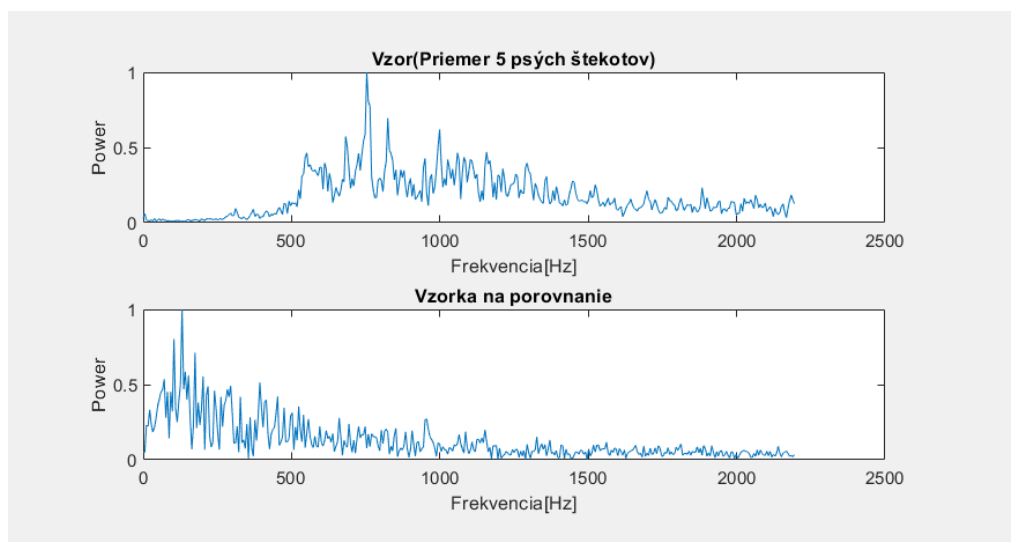
$$C_p = \left| \mathcal{F}^{-1} \left\{ \log \left(|\mathcal{F}\{f(t)\}|^2 \right) \right\} \right|^2 \quad (3.6)$$

ale následné porovnávanie a vykresľovanie neprinášalo chcené výsledky pretože výsledky cepstier sa dosť odlišovali a bolo ťažké určiť nejakú podobnosť.



Obr. 3.5: Vykreslenie spektier pre Vzor a Vzorku

Neskôr, k tomuto postupu bolo prirobené takzvané oknovanie, kde na začiatku bola vložená zvuková nahrávka pozdĺž ktorého bol posúvaný štekot psa a kde nastala primeraná zhoda, ktorá bola určená na základe Threshold, tak tam bolo vyhodnotené, že jedná o psí štekot. Pre určenie bola použitá korelácia, ktorej výstupom sú hodnoty medzi -1 až 1 na základe podobnosti. Na koniec tohto postupu bolo zobrazené v aký čas štekot nastal od začiatku nahrávky. Threshold v tomto prípade bol určený na hodnotu 0.4.



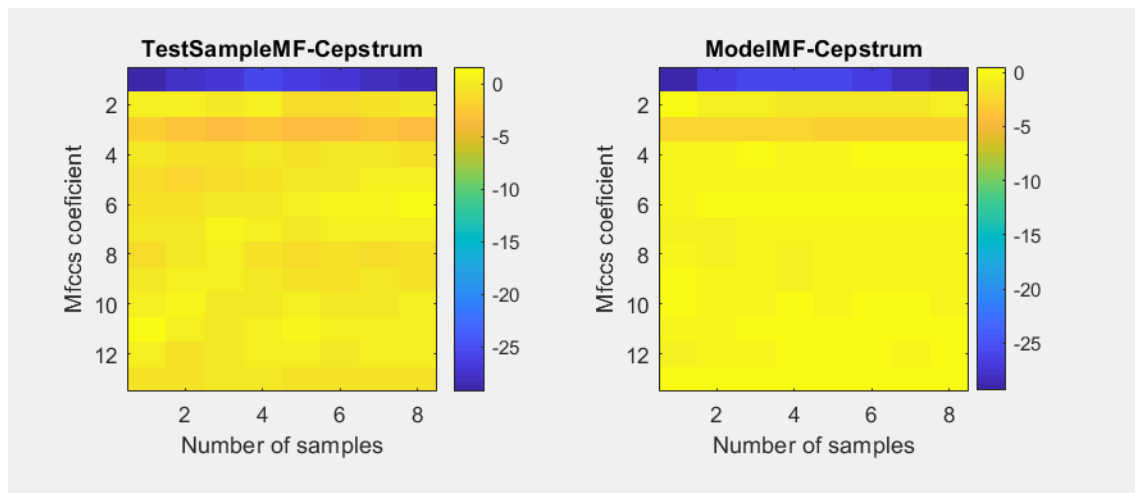
Obr. 3.6: Porovnanie spektra výstrelu a štekotu

Na obrázku 3.6 je vidieť rozdiel spektier medzi spektrom výstrelu a psím štekotom, oba zvuky majú rýchly nábeh amplitúdy čo ich robí podobnými ale prevedením týchto zvukov do spektrálnej oblasti sa dajú jednoducho rozlíšiť.

Tretí postup

Posledný postup ako aj ten druhý využíva príznak z frekvenčnej oblasti. Ide o príznak Mel frekvenčné cepstrálne koeficienty, popis príznaku sa nachádza v prvej podkapitole, postup začína tým že sa vytvorí vzor pre štekot z piatich audio nahrávok z ktorých sa v každom audio segmente extrahuje 13 koeficientov a následne sa vytvorí vzor (jedna matica) ktorá ma pre každý audio segment 13 koeficientov. Takže výsledná matica pre vzor má rozmery 13 krát 8. V ďalšom kroku sa sa spočítajú všetky rady a stĺpce čoho výsledkom je jedno číslo. K tomuto číslu sa porovnáva číslo, ktoré dostaneme po extrakcii Mel frekvenčných koeficientov so vzorky.

Pre porovnávanie bola použitá spočiatku korelácia, ale nakoniec bolo rozhodnuté že sa stanoví Threshold a výsledky budú klasifikované na základe tejto hodnoty.



Obr. 3.7: Vykreslenie pre vzor a vzorku Mel frekvenčných koeficientov

Tak isto ako v druhom postupe tak aj v tomto bol neskôr pozmenený tento algoritmus, bol vytvorený štekot ako vzor, ktorí sa posúval pozdĺž jedného celého užívateľom vybratého súboru wav. Veľkosť okna aj čas o ktorí sa okno posúvalo bolo vopred stanovené. V každom posunutom časovom úseku bolo jedno číslo(súčet riadkov a stĺpcov MFcc) porovnaný so vzorovým a keď hodnota neprekročila hodnotu Threshold, ktorí bol nastavený na číslo 230 tak sa úsek vyhodnotil ako psí štekot a ak prekročil túto hodnotu tak ako iný zvuk. Výsledok bol s čísla úseku prevedený na sekundy a zobrazený.

V grafe 3.7 sú vykreslené Mel frekvenčné cepstrálne koeficienty ako sa menia v každej jednej úseku pre vzor aj pre aj testovaciu vzorku.

Dĺžka štekotu

V praxi sa stretávame s rôznou dĺžkou štekotu, Ide o interval s približnou dĺžkou zhruba od 25ms po 130ms. V tejto práci sa najviac pracovali so štekotom, ktorí mal dĺžku 31 ms. Hranica pod ktorou algoritmy mali problém s detekciou a milným si psieho štekotu je 12ms pod touto hodnotou je už ťažké určiť o aký zvuk sa jedná. Táto hodnota bola určená metódou postupného skracovania štekotu na takú hodnotu kde sa hodnota podobala ostatným zvukom a nedalo sa určiť že sa jedná o psí štekot.

4 Vytvorenie databázy

Obsahom databázy sú výhradne moje nahrávky, ktoré som zozbieral počas niekoľého mesiaca na Slovensku. Zameriaval som sa na rôznorodých psov (starých, mladých, veľkých a malých) Nahrávanie prebiehalo zväčša v exteriéri na mobilný telefón Iphone 6. Snaha bola o čo najbližšie snímanie pri psej čelusti tak isto v čom najpokojnejšom prostredí. Zozbierané nahrávky boli následne prevedené na zvukový formát wav so vzorkovacou frekvenciou 44100Hz s bitovou hĺbkou 16 bitov a bitovou rýchlosťou 1411kbps.

Databáza obsahuje psov Americká akita(10 rokov), Belgický ovčiak(10 rokov), Mudi(6 rokov), Nemecký špic trpasličí(2 roky), Ratlík(8 rokov), Yorkshire terrier (6 mesiacov). Z každého psa boli zozbieraných približne 15 štekotov okrem psa ratlíka, ktorý bol použitý v programe 'Moj pes', u tohoto psa bolo nahratých vyše 60 zaštekání. Ďalej boli jednotlivé psy združené do dvoch kategórií Veľký pes, Malý pes. Každá z kategórií obsahovala 20 nahrávok ktoré v ďalšej časti boli použité k natréňovaniu programu. Je treba spomenúť, že MFCCs pracuje najlepšie pri čo najmenšom hluku, preto u psa Belgický ovčiak, ktorého štekoty boli nahraté v exteriéri, bol použitý HP filter ktoré odrezal spodné frekvencie od 50Hz nižšie.

Názov súboru	Počet súborov	Dĺžka súborov[s]
Americká akita 10 rokov	13	0.32
Belgický ovčiak 10 rokov	14	0.32
Mudi 6 rokov	12	0.32
Nemecký špic trpasličí 2 roky	12	0.32
Ratlík 8 rokov	35	0.32
Moj_pes_testovacie nahrávky(ratlík)	20	0.32
Maly_pes	20	0.32
Velky_pes	20	0.32
Yorkshire Terrier 6 mesiacov	12	0.32
Ľudský štekot	1	12
Moj_pes_test		
-> Moj_pes_test_iba_psy(ratlík)	1	12
-> Moj_pes_test_iba_psy(rôzne psy)	1	17
-> Moj_pes_test_zvuky+psy	1	23
Pes_od_iných_zvukov_test		
-> Psy_test	1	22
-> Zvuky_bez_psa_test	1	83

Tab. 4.1: Prehľad zvukov v databáze

Prehľad o názve súborov, počte súborov v danom priechinku a dĺžke súborov je uvedený v tabuľke V každom súbore s dĺžkou 0.32s je len jeden štekot, avšak v priechinkoch Moj pes test a Pes od iných zvukov test sa nachádza vždy 20 zvukov(buď 20 štekotov 20 iných zvukov alebo ich kombinácia).

Databáza obsahuje aj simulovaný štekot človekom.

5 Vytvorenie uceleného programu na spoľahlivú detekciu

V kapitole 3 boli popísané tri metódy pre detekciu psa spomedzi rôznorodých zvukov, ako som uvádzal najspoľahlivejšia a najperspektívnejšia metóda pre mňa bolo metóda s použitím Mel Frekvenčných Cepstrálnych koeficientov. Táto metóda bola vybraná na finálnu detekciu a tak isto poopravená pre dosiahnutie najlepších výsledkov.

Pred tým ako dôjde k samotnému zaradeniu jednotlivých zvukov dochádza k tzv. trénovaniu sekvencie. Kde z vybraných zvukov z databázy sa vytvára vzor, akýsi priemer koeficientov vybraných zvukov, ktorí je následne použitý v ďalšej časti programu (pri zaraďovaní).

V mojom prípade bolo potrebné vytvoriť 3 vzory (3 vzorové matice) pre veľkého psa, zo psích štekotov Belgického ovčiaka Americkej akity a so psa Mudi, pre malého psa, zo psa ratlík, nemecký špíc trpasličí a psa yorkshire a nakoniec vzor v programe 'Môj pes' ktorí bol vytvorený len so štekotov psa ratlíka.

Spomínané vzorové matice boli uložené do súborov .mat, ktoré obsahovali okrem vzorovej matice aj názov (pomenovanie triedy) ďalej názvy jednotlivých súborov použitých na trénovanie, časové okno a krok z ktorým sa okno posúvalo pri trénovaní.

Každý štekot v trénovanej časti bol upravený na dĺžku 0.32 sekundy čo znamenalo maticu o veľkosti [13,7] 13 koeficientov a 7 časových úsekov. Ku natrénovaniu každého vzoru bolo použitých 20 štekotov, ktoré sa následne nepoužívali v testovacej časti.

-28.1706	-28.1939	-28.3384	-28.6932	-29.1946	-29.8291	-30.5774
-0.7102	-1.5703	-1.8723	-1.7695	-1.6750	-1.2446	-0.9931
-2.1592	-2.3209	-2.2440	-2.3902	-2.7239	-2.5086	-2.2059
0.2838	0.3999	0.6541	0.8094	0.7409	0.7923	0.8313
-0.0406	0.0956	0.1874	0.2313	0.0980	0.1081	0.0515
0.2418	0.2402	0.1625	0.2112	0.2213	0.4226	0.5149
-0.3917	-0.4685	-0.5212	-0.5572	-0.5729	-0.4906	-0.3625
0.3711	0.2171	0.2376	0.1936	0.1825	0.1947	0.3581
-0.3775	-0.2276	0.0295	0.0226	-0.0325	-0.1782	-0.1618
0.1108	0.2471	0.1369	0.0483	-0.0500	-0.1263	-0.1686
-0.0119	0.0730	0.0727	0.2448	0.3047	0.2006	0.0210
0.0865	0.0439	0.0583	-0.1839	-0.0427	-0.0392	-0.1659
-0.0323	-0.0609	-0.0999	-0.0736	-0.0244	-0.0484	-0.0476

Tab. 5.1: Koeficienty vzorovej matice pre malého psa

-28.5866	-28.3990	-28.6250	-28.8064	-29.2308	-29.6383	-31.0312
0.1905	-0.8442	-1.4079	-1.4439	-0.8375	0.2172	1.3276
-2.4630	-2.4304	-2.4353	-2.4339	-2.1758	-2.2848	-2.2372
-0.3228	-0.3496	-0.3613	-0.3391	-0.3074	-0.4476	-0.5085
-0.7355	-0.8940	-0.5894	-0.4196	-0.4395	-0.8891	-1.0776
-0.3262	-0.1547	0.2421	0.3831	0.2382	-0.1392	-0.2142
-0.0772	0.4210	0.7851	0.7819	0.4973	0.4111	0.4274
0.2356	0.8792	0.6647	0.4420	0.3742	0.1663	0.2562
0.3398	0.4644	-0.1477	-0.1879	0.0773	0.1752	0.1978
0.2349	-0.3178	-0.3272	0.0216	0.3126	0.3672	0.1078
-0.0727	-0.2941	-0.0587	0.1322	0.0185	0.1282	0.1924
-0.0687	0.1570	-0.0220	-0.0902	-0.1211	-0.0130	0.1578
0.1793	0.1970	0.1600	0.2505	0.2800	0.2235	0.1425

Tab. 5.2: Koeficienty vzorovej matice pre veľkého psa

V tabuľkách 5.1 a 5.2 sú zobrazené vzory matíc pre malého a veľkého psa.

Na obrázku č. 5.1 je vidieť príklady dvoch rôznych štekotov a ich 13 koeficientov v zvislom rade a 7 časových úsekov.

Po prevedení jednotlivých zvukov na spomínané matice $[13,7]$ sa matice spriemerujú a dostávame výslednú maticu taktiež $[13,7]$ ktorá je vzorom a je uložená. Výsledné matice sú zobrazené v tabuľkách č. 5.2 a č. 5.1

val(:, :, 1) =						
-28.1837	-25.4643	-28.3971	-27.2463	-32.3920	-33.4486	-37.9729
2.2868	-1.3000	-2.3086	-1.3107	0.7180	1.4001	2.5579
-2.6576	-3.3147	-1.9074	-3.0119	-1.9480	-0.7850	-0.2054
0.7203	0.5686	1.7279	1.1081	1.2257	0.8094	1.2681
-0.2618	0.2778	0.6403	0.4067	0.7774	0.6556	0.4835
0.8532	1.5799	0.5685	1.0081	0.8284	0.4514	0.3622
-0.3150	-0.5004	-1.3401	-1.0820	-0.6442	-0.4295	-0.7519
0.3085	-0.0052	0.8893	-0.6365	0.0652	0.2073	0.5972
-0.3632	-0.3531	0.0150	-0.1448	-0.0797	-0.6678	0.3054
0.1421	0.5902	-1.0220	0.0287	-0.0880	-0.1810	0.0450
0.4562	-0.3181	-0.4136	-0.1741	0.0046	-0.3506	-0.0892
-0.6427	-0.5832	-0.2592	-0.1878	-0.2736	0.0187	-0.0510
0.0570	0.2465	0.2465	0.0589	0.0532	-0.0841	0
val(:, :, 2) =						
-25.1771	-26.7418	-27.7714	-28.7264	-32.3998	-35.2579	-40.1532
-0.0236	-2.3038	-1.4494	1.1661	2.2063	2.8856	2.4971
-2.3734	-2.7943	-2.6536	-2.3616	-0.6428	0.1301	0.0677
0.3348	1.2092	1.5113	0.4545	0.4083	0.2252	1.0063
-0.0455	0.1197	0.2771	-1.0590	-0.5430	-0.3451	0.1945
0.0781	0.7129	1.8242	0.5162	0.5367	0.3564	0.4059
-0.1838	-0.8558	-0.7479	-0.4805	-0.2779	-0.3713	-0.3387
0.8475	0.0978	-0.8165	-0.1997	-0.0684	-0.3483	-0.2302
-0.8103	-0.1932	0.1624	-0.1380	0.3041	-0.0922	0.0819
-0.0589	0.9531	0.5654	0.0785	0.0539	-0.1427	0.1730
-0.4154	-0.2752	-0.3648	-0.0285	-0.5693	-0.3044	-0.5529
0.3561	0.0997	0.0608	-0.1034	-0.2845	-0.2651	-0.2648
-0.1065	-0.2595	-0.4190	-0.2595	-0.1209	-0.0312	0

Obr. 5.1: Príklady dvoch rôznych štekotov a ich koeficienty

5.1 Testovanie

Na začiatku testovacej časti boli načítané jednotlivé natrénované vzory pre veľkého a malého psa. Dôvod tréningu dvoch vzorov bol ten, že koeficienty veľkého a malého psa sa niekedy dosť líšia a použitie len jedného vzoru vedie k častejším preklepom.

Program pracuje s tzv. oknovaním, kde testovací zvuk sa rozseká na malé úseky z ktorých su následne vyrátané MFCCs a ktoré sú následne porovnané s MFCCs vzoru. Porovnanie je zrealizované tak že všetky koeficienty z každého úseku sa postupne odčítajú, následne umocnia a na koniec sa zrátajú všetky stĺpce a riadky výsledkom čoho je jedno číslo. Ukážka na obrázku č.5.2

```

d{i}= ((Curframe{i}-mfccFeatures{1}).^2);
d{i} = sum(d{i});
d{i} = sum(d{i},2);

```

Obr. 5.2: Porovnanie koeficientov

$$D = \sum_{i=1}^{+13} \sum_{j=1}^{+7} [S_{ij} - Z_{ij}]^2 \quad (5.1)$$

Vzorec 5.1 je matematickým vyjadrením porovnávania koeficientov. Symbol D predstavuje výsledné číslo, symbol S súčasný zvukový segment a Z vzor (vzorová matica).

Následne boli stanovené tzv. 'Threshold' kde hodnota nad stanovené číslo znamenala 0 vo finálnej matici, kdežto keď bola hodnota identifikovaná pod stanovenú Threshold do matice bola zapísaná 1.

Ako je uvedené na obrázku 5.2 výsledkom daného výpočtu dostávame jedno číslo, ktoré ako je uvedené vyššie sa porovnáva proti dvom 'Threshold', tým pádom dostávame 2 finálne matice, ktoré obe majú jeden riadok rovnaký počet stĺpcov a s ktorými sa ďalej pracuje.

Na koniec medzi dvoma poslednými maticami sa vykonala operácia OR po ktorej dostávame poslednú maticu kde jednotky znamenajú, že v danom úseku bol identifikovaný psí štekot a nuly, že sa psí štekot v danom úseku nenachádza. Nakoniec sú výsledky zobrazené spôsobom znázorneným na obrázku 5.3

Pre zlepšenie detekcie, okrem používania rôznych vzorov pre dvoch odlišných psov bolo použité odstránenie niektorých koeficientov. Odstránené boli také koeficienty ktoré vykazovali také isté, alebo veľmi podobné hodnoty pre dva rôzne zvuky. Napríklad štekot a hovorené slovo. Týmto spôsobom sa docielilo to že, koeficienty ktoré sa odlišovali mali na výsledné číslo, po sčítaní riadkov a stĺpcov ako je ukázané na obrázku 5.2 väčší dopad.

Odstránenie koeficientov bolo prevedené na základe metódy pokus omyl a niektoré varianty, aj s ich percentuálnou úspešnosťou sú zobrazené v tabuľke ??

5.2 Určenie Threshold

Určenie Threshold alebo tzv. hodnote, ktorá rozhoduje medzi tým či testovaný zvukový signál bude zaradený ako psí štekot alebo nie, bolo zrealizované metódou pokus

```
Dog detected at second 1.35
Dog detected at second 1.45
Dog detected at second 1.55
Dog detected at second 3.45
Dog detected at second 3.55
Dog detected at second 3.65
Dog detected at second 9.55
```

Obr. 5.3: Výsledné zobrazenie

omyl. Otestoval som väčšinu štekotov z databázi a zapisoval som si približne, aké hodnoty vykazuje ich rozdiel od vzoru a následne som stanovil približnú hodnotu, Threshold.

Pre zvuky iné ako psí štekot som postupoval tak isto a niektoré hodnoty skúšané aj s pravdepodobnosťou sú zobrazené v tabuľke 5.3.

Ako je zrejmé z tabuľky 5.3, presné a konečné stanovenie Threshold bolo dosiahnuté postupným menením hodnôt Threshold pre dosiahnutie najlepších hodnôt percentuálnej úspešnosti TPR a TNR. Na koniec sa ponechali hodnoty ktoré sú v tabuľke označené ako Original values.

5.3 Vyhodnotenie

Ako spôsob vyhodnotenia činnosti algoritmu bola použitá tzv. Confusion matrix. Kde každý z testovaného signálu bol zaradený do jednej so 4 možností, TP(True positive), TN(True negative), FP(False positive), FN(False Negative). Zaradenie do skupiny TP by znamenalo to že zvuk(psí štekot) bol klasifikovaný správne a zaradenie do skupiny TN by znamenalo, že nesprávne. Takto isto sa klasifikovali zvuky odlišné od psieho štekotu. Po roztriedení zvukov na štyri časti, ako je uvedené vyššie, prichádza na radu výpočet percentuálnej úspešnosti algoritmu, výsledkom čoho sú dve čísla vyjadrené v percentách. True positive rate 5.2 a True negative rate 5.3. Výsledky percentuálnych úspešností a všetky spomenuté kategórie su zhrnuté a zapísané v tabuľke 5.3.

	Použité všetky koeficienty							
	Threshold		TP	TN	FP	FN	TPR[%]	TNR[%]
	Malý pes	Veľký pes						
Original values	190	165	18	16	4	2	90	80
	170	155	16	15	5	4	80	75
	145	130	13	18	2	7	65	90
	210	175	20	10	10	0	100	50
	Prvý koeficient eliminovaný							
	Threshold		TP	TN	FP	FN	TPR[%]	TNR[%]
	Malý pes	Veľký pes						
Original values	91	63	18	15	5	2	90	75
	100	80	19	13	7	1	95	65
	115	90	20	10	10	0	100	50
	80	50	15	16	4	5	75	80
	Ponechané prvé 3 koeficienty							
	Threshold		TP	TN	FP	FN	TPR[%]	TNR[%]
	Malý pes	Veľký pes						
Original values	60	120	18	17	3	2	90	85%
	45	100	14	18	2	6	70	90
	35	65	10	19	1	10	50	95
	70	150	20	13	7	0	100	65

Tab. 5.3: Výsledná pravdepodobnosť, hodnoty Threshold a počet ponechaných koeficientov

$$TPR = \frac{TP}{TP + FN} * 100 \quad (5.2)$$

$$TNR = \frac{TN}{TN + FP} * 100 \quad (5.3)$$

5.4 Program Môj Pes

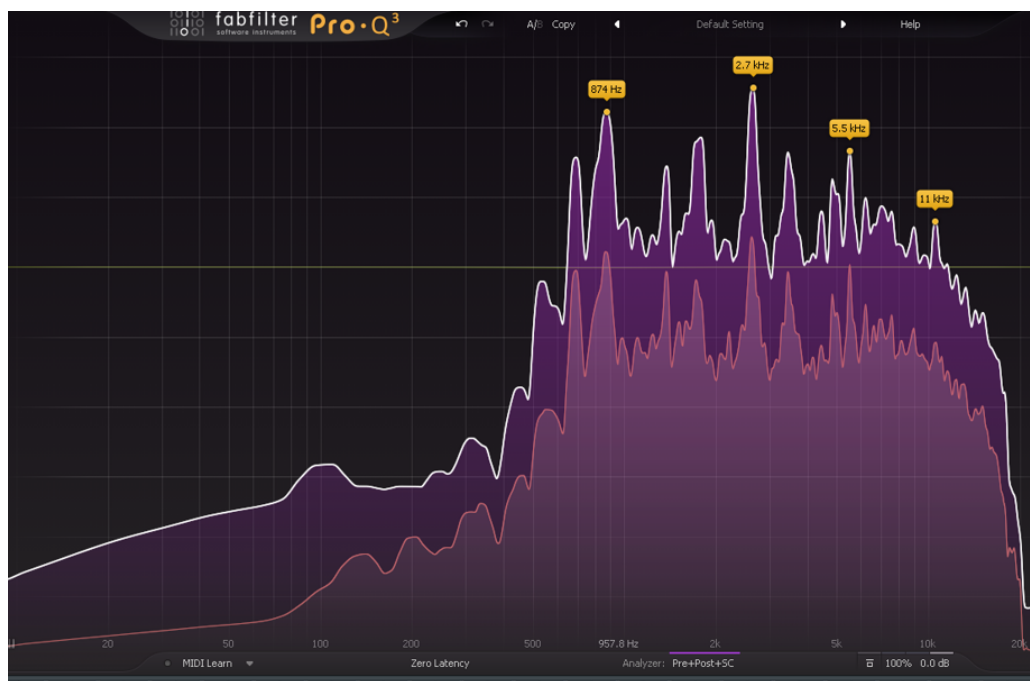
V programe s názvom Moj Pes bol hlavný cieľ rozlíšiť jedného psa od ostatných psov ako aj od iných zvukov. V tomto postupe bola použitá rovnaká metóda ako v predchádzajúcom prípade, len trochu po-upravená. Na rozdiel od všeobecnej metódy, kde sa rozlišovalo medzi psom a nie psom a kde sa ponechali prvé 3 koeficienty, v programe Moj pes boli ponechané všetky koeficienty okrem prvého v poradí. Postupom pokus omyl sa určil Threshold tak ako v predchádzajúcom prípade. Hodnota Threshold sa stanovila na hodnotu 45 a percentuálna úspešnosť detekcia bola určená TPR 85%,

TNR 95% (Rozlíšenie Moj pes od rozličných zvukov). a TNR 85% (Rozlíšenie Moj pes od iných psov).

Iné zvuky rozpoznané ako štekot

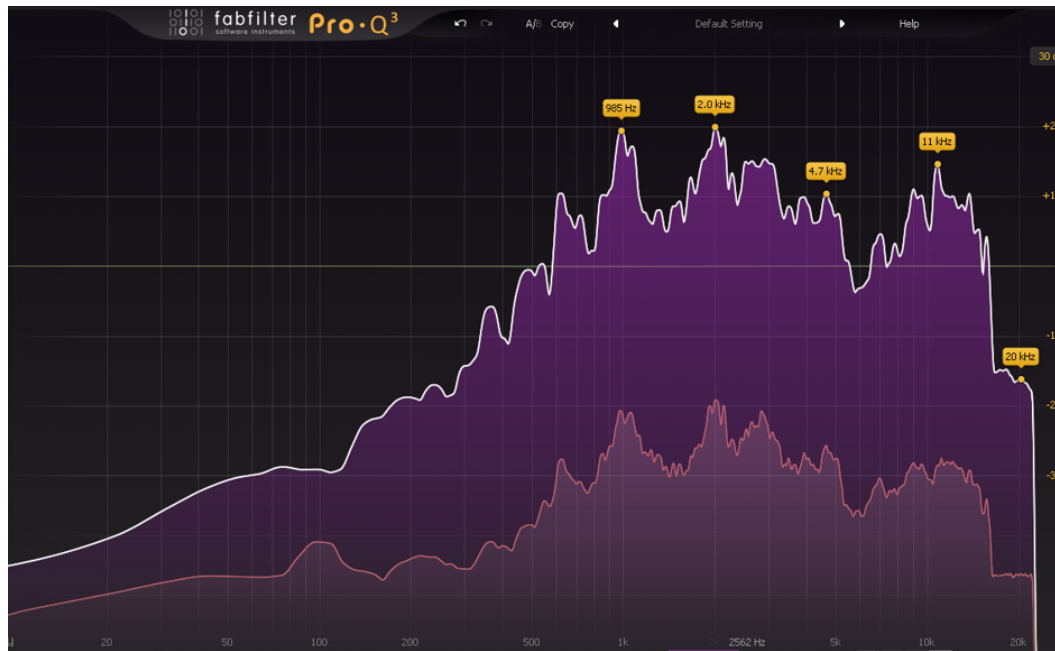
Zvuky, v testovacej sekvencii, ktoré boli rozpoznané ako pes sú zvuky vydané sliep-kou, agresívne mňaučanie mačky, a smrkanie do vreckovky, ostatné zvuky v sekvencii boli určené správne. Tieto zvuky sa v nahrávke vyskytovali každý len jeden krát. V programe Moj Pes boli chybne určené dve zaštekania psa Americká akita a jedno šteknutie od Belgického ovčiaka. Štekot americkej akity bol použitý päť krát v sekvencií a štekot Belgického ovčiaka 6 krát.

Na obrázkoch 5.4 a 5.5 sú príklady dvoch spektier, psieho štekotu od malého psa a zvuku vydávaného sliepkou. Zvuk vydávaný sliepkou v sekvencii bol určený ako psí štekot tým pádom nesprávne. Dôvodom môže byť veľmi podobné spektrum, ktorej dominantné zložky sú si tak isto veľmi podobné. Zvuk vydaný sliepkou ma dominantné zložky na frekvenciách 874Hz 2.7kHz 5.5kHz a 11kHz. Psí štekot má dominantné zložky na frekvenciách 985Hz 2kHz 4.7kHz 11khz. Ostatné zvuky určené nesprávne majú spoločné to isté, ich frekvenčné spektrum a ich spektrálne zložky sú si podobné. Taktiež si môžeme všimnúť, že jednotlivé spektrá nemajú veľké



Obr. 5.4: Spektrum zvuku vydávané sliepky

zastúpenie na spodných frekvenciách okolo 100-200Hz čo podobnosť medzi zvukmi ešte zvyšuje. Ďalší podobný prvok medzi týmito dvoma zvukmi je rýchly nábeh amplitúdy, avšak tento jav nemá na výslednú podobnosť vplyv, pretože MFCCs pracujú v spektrálnej oblasti.



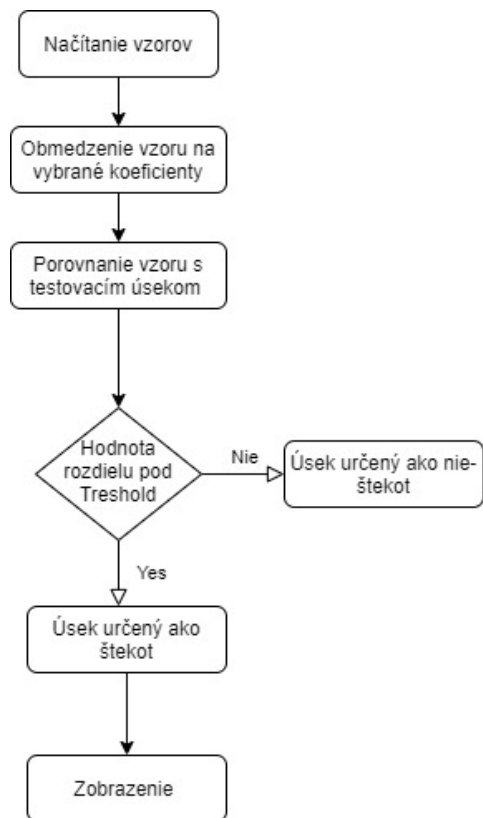
Obr. 5.5: Spektrum psieho štekotu

Zvuky určené sprave, ako nie pes sú zvuky s rozdielnym spektrom aj fonematickou štruktúrou popis funkcie MFCC je v chapter 3.

5.5 Princíp a inštrukcie ku práci s programom

Princíp použitých algoritmov je popísaný v predchádzajúcich paragrafom ale prehľadná zjednodušená vizualizácia je zobrazené na obrázku č.5.6

Použitie a testovanie jednotlivých zvukových nahrávok sa odohráva v prostredí matlab, pre optimálnu činnosť boli vytvorené viaceré funkcie ale pre testovanie a prípadného užívateľa sú potrebné len dve. Funkcia s názvom calling-my-dog ktorá bola použitá pre program Moj pes a funkcia callingmfs, ktorá bola vytvorená pre program rozpoznania psa od iných zvukov.



Obr. 5.6: Zjednodušený stavový diagram pre Rozpoznávanie psích štekotov

Používateľ vkladá do funkcie wav súbor a rôznej dĺžky, ktorý je následne vyhodnotený a výsledky sú zobrazené.

Záver

Cielom semestrálnej práce bolo detekovať štekot psov spomedzi iných zvukov v časovej aj frekvenčnej oblasti použitím troch algoritmov. Ciel práce bol splnený vytvorili sa tri postupy prvý používa príznaky s časovej oblasti zero-crossing rate energiu a energiu entropiu. Postupne sa pomocou vzdialenosti porovnávali rozostupy medzi jednotlivými audio segmentami a na základne vo pred stanovenej hodnoty Threshold sa určilo či sa jedná o psí štekot alebo nie. V tejto prvej metóde bolo veľmi obtiažne určiť správne hodnoty Threshold pretože výsledky príznakov skoro vždy vychádzali malé hodnoty ktoré sa často prekrývali a bolo náročné určiť presnú hranicu kedy sa jedná o psí štekot a kedy nie. Úspešnosť metódy, že detekuje psa aj že určí v akej sekunde bolo psí štekot nájdený bola mála okolo 60%. Výsledok pripisujem tomu, že po extrakcii troch príznakov ktoré vykazovali veľmi malé hodnoty a tomu že často dvom rozdielnym zvukom sa pripisovali veľmi podobné hodnoty. V pokračovaní práce by som rozmýšľal nad obmenou príznakov.

Druhá metóda pracovala na podobnosti spektra vo frekvenčnej oblasti. Psí štekot sa porovnával s nahrávkou dĺžky 15 sekúnd a keď nastala vo pred určená zhoda výstup kódu určil, kedy bol daný úsek v 15 sekundovej nahrávke rozpoznaný ako psí štekot. Jednotlivé malé úseky sa porovnávali pomocou korelácie. Úspešnosť detekcie psieho štekotu bol 70% Postup s porovnávaním spektier rieši problém z prvého postupu kde ako príznaky sa používala energia a energia entropia a zero-crossing-rate. Energie v audio segmente pre psí štekot aj vystrel zo zbrane sú veľmi podobné pretože u obidvoch prípadov ide o náhly a výrazný zvuk ale keď je použitý príznak spektrálne koeficienty tento problém sa vyrieši.

Posledná metóda pracuje s príznakom Mel Cepstrálne koeficienty. Táto metóda podobne ako aj tá druhá postupne porovnáva a posúva psí štekot pozdĺž 15 sekundového úseku kde ako výstup je matica ktorá ma 13x8. Tieto hodnoty v matici sú následne sčítané a výsledky sa porovnávajú s Threshold. Ide o najpresnejší postup pretože sa spriemeruje veľa hodnôt a určovanie Threshold nebolo tak náročné. Úpešnosť metódy je 75 %.

Určená minimálny dĺžka signalu pre správnu detekciu bola 12ms pod touto hranicou je už obtiažne určiť o aký zvuk sa jedná.

V pokračovaní práce by som sa rád zamerail na Mel Cepstrálne koeficienty pretože podľa testov vykazovali najlepšie výsledky rád by som ich skombinoval s prvým postupom pretože v ňom nastávali dosť časté preklepli.

V pokračovaní v letnom semestri som naviazal na poznatky nadobudnuté v zimnom a použil som jedno z metód skôr odskúšanú. Cieľom bolo viac a boli splnené nasledujúcim spôsobom. Na začiatku bolo potrebné vytvoriť databázu viacerých štekotov, ktoré obsahovala rôznorodých psov a ktorej obsah je prehľadne uvedený

v tabuľke 4.1. Neskôr databáza bola doplnená o simulovaný štekot človekom. V nasledujúcom kroku a pomocou zostavenej databázy boli vytvorený program na spoľahlivú detekciu psa spomedzi iných zvukov. Základom tejto metódy bolo použitie MFCCs ktorej podrobný popis je v chapter 5. Výsledky sú zaznamenané v tabuľke 5.3 kde najlepší výsledok bol zaznamenaný pre ponechané prvé tri koeficienty, úspešnosť TRP 90% a TNR 85%.

Ďalšou časťou zadanie bolo vytvorenie programu, ktorí som nazval Moj pes, podstatou programu bolo odlíšiť štekot jedného jedinca od ostatných štekotov a aj od ostatných zvukov. Úspešnosť detekcie TPR 85% TNR 95%(Pes od psov) a TNR 85%(Pes od iných zvukov). V tomto programe boli ponechané všetky koeficienty okrem prvého v poradí.

Zvuky ktoré program rozpoznal ako štekot a možný dôvod chybného určenia je podrobnejšie popísaný v sekcii section 5.4.

Prácu hodnotím v pozitívnom duchu a pri budúcej práci s podobnou úlohou by som sa rád zamerlal na čo najkvalitnejšie nahrávky v akusticky upravenom prostredí v interiéri. Keďže som v bakalárskej práci používal výhradne svoje nahrávky(v semestrálnej práci nahrávky boli stiahnuté z internetu) a pri súčasnej situácii nahrávky neboli najkvalitnejšie. MFCCs najlepšie pracuje pri čo najmenšom šume, táto skutočnosť mohla mať dopad na výsledky ktoré mohli byť okolo 95%.

Literatura

- [1] GIANNAKOPOULOS, Theodoros a Aggelos PIKRAKIS. Introduction to audio analysis: a MATLAB approach. Oxford: Academic Press, 2014. ISBN 978-0-08-099388-1.
- [2] VERALDO, Valerio. Waveform [online]. In: . 29,6,2020 [cit. 2020-11-19]. Dostupné z: <<https://www.youtube.com/watch?v=bnHHVo3j124t=1146s>>
- [3] TÖRÖK, Jozef. Návrh a výroba extruzního stroje pro výrobu tiskové struny [online]. Brno, 2020 [cit. 2020-11-30]. Dostupné z: <http://hdl.handle.net/11012/190516>. Bakalářská práce. Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Ústav automatizace a měřicí techniky. Vedoucí práce Ondřej Baštán.
- [4] Che Yong Yeo, S. A. R. Al-Haddad and Chee Kyun Ng, "Dog voice identification (ID) for detection system,"2012 Second International Conference on Digital Information Processing and Communications (ICDIPC), Klaipeda City, 2012, pp. 120-123, doi: 10.1109/ICDIPC.2012.6257264.
- [5] INICALO. GTZAN music/speech collection [online]. 12.10.2017 [cit. 2020-12-13]. Dostupné z: <https://www.kaggle.com/lnicalo/gtzan-musicspeech-collection>
- [6] YIN, Sophia a Brenda MCCOWEN. Barking in domestic dogs: context specificity and individual identification [online]. Department of Animal Science, University of California, Davis, USA, 21 May 2004n. l. [cit. 2020-12-13].
- [7] KOMORI, Yuichi, Kazuaki OHNO, Takuaki FUJIEDA, Suzuki TAKAHIRO a Tadokoro SATOSHI. Detection of continuous barking actions from search and rescue dogs' activities data [online]. IEEE, 17 December 2015 [cit. 2020-12-13].
- [8] SUZUKI, Matsumi. Apparatus for determining dog's emotions by vocal analysis of barking sounds [online]. 13 June 2004 [cit. 2020-12-13].
- [9] SIGMUND, Milan. A Tool for Automatic Keyword Search in Audio Documents [online]. Brno University of Technology, Faculty of Electrical Engineering and Communication, 24 January 2012n. l. , 4 [cit. 2020-12-13].
- [10] JAN, J. Číslicová filtrace, analýza a restaurace signálů. Brno: VUT v Brně, 2002.
- [11] TIMS, K. Zero crossing. Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001-, 1. 4. 2006 [cit. 2021-4-27].

- [12] ILYIN, D. Signal sampling representation. Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001-, 6. 1. 2021 [cit. 2021-4-27].

Seznam symbolů, veličin a zkratk

DSP	Číslicové zpracování signálů – Digital Signal Processing
MFCCs	Mel-frequency cepstral coefficients – Mel-frekvenčné cepstrálne koeficienty